Course	**COMPUTATIONAL STATISTICS FOR POLITICAL SCIENCE**			
	Type	Semester	ECTS	Code
	(E)	III	4	
Course instructor **Course assistant** **Course tutor**				
Course objectives and learning outcomes	Our century is the century of science, and science is a social endeavour that produces knowledge on empirical data. There is no area of human activity today that does where scientific knowledge does not build on data. As a discipline within the larger social sciences, Political Science does not differ in this aspect. If we want to bring meaning in the word "science" on the name of our discipline, then collecting and analysing data should be our main activities and their study should be an important component of our program. Today Statistics and Data Science, like every other research activity have been heavily computerized. The dramatic increase of available data, especially during the second half of the XX-th Century and on has brought about the need to master computer programs and programin languages that would perform calculations of the large amount of data that we possess. Those skills are a must not only in the academic research activity, but also in a professional life as public servants or public policy drafters and implementers. Most of the quality job calls today in our country and around the world ask for candidates with knowledge in statistical packages and programming languages. This course aims just that: introducing students in the world of statistical analysis with the statistical package Stata and the programming language R. Its objective is that at the end of the course, students could confidentially write in their CV-s a decent knowledge of at least one of those programs and a general understanding of the other. This course aims at strengthening the skills component of our program: our students should not only know stuff but also know to make stuff. Moreover, this course represents one more component of our orientation toward research-oriented teaching, which would allow students more independence in the data manipulation and analysing process. A second outcome of this course is the strengthening of students' knowledge acquired in two other important courses of our program, Research Methods 1, and Research Methods 2. Although this course is rooted in statistics, our examples will come from social and behavioural sciences, especially Political Science. We will use data that other researchers have gathered as well as our own data that we have collected over the years through practicums with collecting data through public opinion surveys.			
Learning outcomes	Upon successful completion of this course, students will be able to: • The capability of practically applying knowledge acquired in the Research Methods 1 (ResearchDesign), and Research Methods 2 (Statistics) courses; • Mastering the most important and practical elements of the statistical packages Stata and the Rprogramming language; • Acquiring the essentials of programming for data science; • Student preparation to write publishable diploma thesis by using big data and statistical analysis			
Content	**Weekly program**			**Week**

	1. STATA AND ITS SOURCES Typography; Stata environments; Stata auxiliary documentations, searching information; Statalist, *Stata Jounal*; literature	1
	2. INTRODUCTION TO R AND ITS SOURCES The R editor environment and R Studio; the RData format; R auxiliary documents, the Rblog, *The R Journal*	
	3. MENAGING DATA IN STATA AND R Command examples; creating a new dataset; sub-dataset specification – the inandifqualifiers; creating and replacing variables; using functions; converting between numeric and string variables; creating new categorical and ordered variables.	2
	4. MENAGING DATA IN STATA AND R Import data from other variables; combining two or more Stata documents; transposing, reformatting and compressing data; weighting observations; creating random data and samples; writing programs for data management; memory management.	3
	5. GRAPHS IN STATA AND R Command examples; histograms; scatterplots; line plots; connected line plots; other twoway plot types; box plots; pie charts; bar charts; dot plots;symmetry and quantile plots; quality control graphs; adding text to graphs; overlaying multiple twoway plots; graphing with do-files; retrieving and combining graphs.	4
	6. STATISTICAL ANALYSIS AND TABLES IN STATA AND R Command examples; summary statistics for measurement variables; exploratory data analysis; normality test and transformations; frequency tables and two-way cross-tabulation; multiple tables and multi-way cross-tabulation; table of means, medians, and other summary statistics; using frequency weights.	5
	7. ANOVA AND OTHER COMPARISON METHODS IN STATA AND R Command examples; one-sample test; two-sample tests; one-way analysis (ANOVA); two- and N-way analysis of variance; analysis of covariance (ANCOVA); predicted values and error-bar charts	6
	8. LINEAR REGRESSION WITH STATA AND R Command examples; the regression table; multiple regression; predicted values and residuals; basic graphs for regression; correlations; hypothesis tests; dummy variables; automatic categorical variable indicators and interactions; stepwise regression; polynomial regression; panel data.	7
	9. REGRESSION DIAGNOSTICS WITH STATA AND R Command examples; SAT score regression, revisited; diagnostic plots; diagnostic case statistics; multicollinearity.	8
	10. FFITING CURVES AND ROBUST REGRESSION WITH STATA AND R Command examples; band regression; lowess smoothing; regression with transformed variables; conditional effect plots; nonlinear regression; regression with ideal data; Y outliners; X outliners; assymetric error distribution; robust analysis of variance; rregandqregapplications;robust estimates of variance.	9

	11. LOGISTIC REGRESSION WITH STATA AND R Command examples; space shuttle data; using logistic regression; conditional effect plots; diagnostic statistics and plots; logistic regression with ordered-category y; multinominal logistic regression.	10
	12. TIME SERIES ANALYSIS WITH STATA AND R Command examples; smoothing; further time plot examples; lags, leads, and differences; correlograms; ARIMA models.	11
	13. INTRODUCTION TO PROGRAMMING WITH THE ADO LANGUAGE Basic concepts and tools; example program: moving autocorrelation; ado-file; help file; matrix algebra; bootstrapping; Monte Carlo simulation	12
	14. INTRODUCTION TO PROGRAMMING WITH THE R LANGUAGE Basic concepts and tools; example program: moving autocorrelation; help file; matrix algebra; bootstrapping; Monte Carlo simulation	15

Teaching methods	**Academic activity**	**Weight (%)**
	- Lectures and seminars	45%
	- Lab work	45%
	- Practicum (optional and outside the academic timeline)	10%

Academic obligations	**Academic obligations**	**Number**	**Week**	**Weight**
	- Homework and classwork	10	2, 3, 4, 5, 6, 7, 8, 9, 10, 11	90% 10%

Sources and concretisation tools	**Tools**	**Numri**
	- Classroom (e.g)	1
	- Computer lab (e.g)	1
	- Moodle	
	- Software Stata, R	2
	- Projector	1

Activity and load	**Activity type**	**Orë javore**	**Ngarkesa**
	1. Lectures and seminars	2	30
	2. Lab work		15
	3. Independent learning		40
	4. Homework and classwork		15

Literature/references	1. Lawrence Hamilton. 2006. *Statistics with STATA*. Belmont, CA: Thomson Books/Cole. 2. Ridvan Peshkopia. 2018. Statistika Kompjuterike me STATA: Përmbledhje Ligjëratash. Prishtinë: UBT. 3. Eralda Dhamo. 2018. Hyrje në R: Përmbledhje leksionesh.
Contact	